DOCUMENT RESUME

ED 165 547 HE 010 774

AUTHOR Sawyer, Richard L.; Maxey, James
TITLE The Validity of College Grade Prediction Equations Over Time.
PUB DATE 27 Oct 78
NOTE 19p.; Paper presented at the Southern Association for Institutional Research, (Nashville, Tennessee, October 26-27, 1978)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Academic Ability; College Admission; *College Bound Students; *College Freshmen; Grade Point Average; *Grade Prediction; Grades (Scholastic); Higher Education; Institutional Research; *Measurement Techniques; Post High School Guidance; Predictive Measurement; *Predictive Validity; Research Projects; Scoring Formulas; Standardized Tests; *Student Evaluation; Test Interpretation
IDENTIFIERS *American College Testing Assessment Program

ABSTRACT

A sample of 260 colleges was surveyed during the years 1972-1976 to determine the validity of predicting college freshmen grades from standardized test scores and high school grades using the American College Testing (ACT) Assessment Program, an evaluative and placement service for students and educators involved in the transition from high school to college. The ACT Assessment consists of four academic tests, self-reported high school grades, the Student Profile Section (SPS), and the ACT Interest Inventory. The academic tests cover the traditional subject areas of English, mathematics, social studies, and natural sciences. A composite score is tallied from these tests and this score is considered a measure of overall academic ability. The SPS provides information on student's background, extracurricular accomplishments, special academic needs, housing plans, financial need, planned major, and career plans. The Interest Inventory measures preferences for job-related activities in six categories: science, creative arts, social service, business contact, business detail, and technical skills. ACT offers two plans to colleges for predicting freshmen grades: (1) the Basic Research Service will predict grades for specific courses and subgroups of students, and can utilize data other than ACT test scores. The quality of grade predictions in this study was stable over the four-year period. Freshmen grade point averages appear to be predictable using equations as old as four years. To increase the accuracy of the predictions it is possible for colleges with small student populations to pool data from two or more years when developing prediction equations. (BH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED165547

S A I R

THE VALIDITY OF COLLEGE GRADE PREDICTION EQUATIONS OVER TIME

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Southern Assn for Instit Research

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Presented to the
Southern Association for Institutional Research
Nashville, Tennessee
October 26-27, 1978

Richard L. Sawyer
James Maxey
Research and Development Division

The American College Testing Program
P. O. Box 168
Iowa City, Iowa

HE 010 774

The Validity of College Grade Prediction Equations Over Time[1]

Richard Sawyer[2] and James Maxey[2]

Introduction

A problem commonly encountered in predicting college freshman grades from standardized test scores and high school grades is the validity of the prediction equations over time. Changes over time in students' academic abilities, in freshman curricula, and in instructors' grading policies can be large enough to make old prediction equations inaccurate. This has obvious negative implications for both colleges and students. On the other hand, collecting and reporting the data needed to revise prediction equations can be expensive and time-consuming. Therefore, many college researchers would naturally want to use prediction equations as long as they could before updating them.

The purpose of this study is to examine the validity over time of prediction equations based on ACT Assessment test scores and high school grades. Prediction equations were calculated using data from the years 1972, 1973, 1974, and 1975. These prediction equations were then cross-validated against freshman grades earned in 1976. The discrepancies between the actual grades and the predicted grades were then compared.

The ACT Assessment Program

The ACT Assessment Program is a comprehensive evaluative, guidance, and placement service for students and educators involved in the transition from high school to college. It is based on the ACT Assessment, which consists of four academic tests, self-reported high school grades, the Student Profile

---

[1]Presented at the 1978 annual meeting of the Southern Association for Institutional Research.

[2]Institutional Services Research and Development Division, The American College Testing Program, Iowa City, Iowa 52240.

Section (SPS), and the ACT Interest Inventory. The program also includes a number of supporting research services.

The four academic tests of the ACT Assessment measure abilities in the subject areas traditionally identified with college and high school programs: English, mathematics, social studies, and natural sciences. The English test measures students' understanding and use of the basic elements of correct and effective writing; the mathematics test, their mathematical reasoning and problem-solving ability; the social studies test, the problem-solving skills required in the social studies; the natural science test, the critical reasoning and problem-solving skills required in the natural sciences. The arithmetic average of the scores on these four tests is the ACT Composite score, which is often used as a measure of overall academic ability.

ACT test scores are reported on a standard scale that ranges from 1 to 36. On this scale, a student's true score is usually within two score points on either side of the score reported for each of the four tests and within one point on either side of the ACT Composite score.

When students register for the ACT Assessment, they report the last grade received prior to the senior year of high school in each of the four subject areas. The arithmetic average of these four grades--the high school average--provides another measure of overall academic ability. The high school average also reflects students' academic motivation and other college grade-related factors. It has been found to be about as predictive of college grades as the commonly used high school rank. For detailed descriptive and technical information about ACT test scores and high school grades, consult the *Technical Report of the ACT Assessment Program* (ACT, 1973).

Another component of the ACT Assessment is the Student Profile Section (SPS). Through the SPS, students provide information about their background, extra-curricular accomplishments, special academic needs, housing plans, financial

need, planned major, and career plans. Students complete the SPS when they register for the ACT Assessment.

The ACT Interest Inventory measures students' preferences for job-related activities in six basic interest dimensions: Science, Creative Arts, Social Service, Business Contact, Business Detail, and Technical. Students complete the Interest Inventory when they register for the ACT Assessment.

Information from all these sources--the ACT tests, high school grades, SPS, and Interest Inventory--is organized into individual Student Profile Reports sent to the students and colleges. With the summary information provided by the Student Profile Report, students and educators can make informed plans and decisions.

The ACT Predictive Research Services

ACT offers without charge to colleges two general plans for predicting freshman grades. Each plan is designed to meet the varying needs and resources of colleges which use ACT data.

The Basic Research Service entails minimal efforts by a college in reporting data. To participate in Basic, college personnel simply mark on a computer-generated roster the overall grade point averages of their freshmen. A minimum data base of 100 records is required. Through the Basic Research Service, a college can obtain predictions of overall freshman grade point average using the four ACT test scores and the high school average. Providing they have data for at least 100 students of each sex, colleges can also obtain separate prediction equations for males and females. During the 1977-78 academic year, 460 colleges reported grades for 126,880 freshmen through the Basic Research Service.

The Standard Research Service is designed for colleges that either want to predict specific course grades, or develop prediction equations for

subgroups of students, or want to use predictors other than ACT test scores and high school grades. A large variety of research studies can be accommodated by the Standard Research Service. Data can be supplied either through punched cards, optically scanned cards, or magnetic tape. During 1977-78, 187 colleges reported grades through Standard Research Service.

Many college applicants take the ACT Assessment in October of the year before they matriculate. Because it is desirable to use the same prediction equation for as many of a college's applicants as possible, ACT does not update prediction weights until October of the year in which students matriculate. As a result, prediction equations are developed from data that are at least two years older than the grades being predicted. ACT requires colleges to participate in Basic or Standard at least once every three years, however, if they want to continue to receive grade predictions for their prospective freshmen. This study was designed to examine the validity of this policy and to recommend any adjustments that seem warranted.

Data Base and Sample Design

The institutions selected for this study are a probability sample of the institutions that participated in the Standard or Basic Research Service in the academic year 1976-77 and in one or more of the academic years 1972-73, 1973-74, 1974-75, and 1975-76. The requirement that an institution participate for two or more years was made to permit cross-validating grade predictions calculated from one year's data on grades earned during 1976.

To maximize the base from which data could be sampled, the variables used in the predictions were restricted to the Basic specifications: overall

grade point average as criterion, high school grades and ACT scores as predictors, and subgrouping on sex.

Most of the overall grade point averages in this study are for first semester freshmen in the fall of 1972, 1973, 1974, 1975, and 1976. Colleges participating in Standard do have the option of reporting second semester freshman grades, and participants in both plans have the option of reporting grades from previous years. In the vast majority of cases, however, colleges report first semester freshman grades for the current academic year.

Because the data in this study are from colleges participating in ACT's predictive research services, they are in two key respects not representative of colleges nationally:

- Colleges using the ACT Assessment are located mainly in the South and Midwest with comparatively fewer in the East/Northeast and West Coast.
- Participation in Standard and Basic is voluntary, so that the sample is self-selected even among colleges that use the ACT Assessment.

The results of the study are therefore not necessarily representative of the results that would be obtained if data from all colleges in the nation could somehow be collected. Readers connected with non-ACT participating colleges should, therefore, be cautious in applying the results to their own prediction problems. Nevertheless, this study should be useful to these readers in suggesting ways of formulating and solving prediction problems and in pointing out major trends.

Stratification Variables

It was anticipated that the efficiency of grade predictions might vary according to differences in the composition and operation of the colleges.

By taking note of these differences in the sampling, greater precision in the inferences from the sample may be obtained. In this survey, separate samples were selected from strata of colleges determined by their control, type, and size. The stratification variables are:

- The control of a college is either public or private.
- The type of a college is determined by the highest degree level it offers:

  Type 1: Two but less than four years of work beyond Grade 12--includes junior colleges, technical institutes, normal schools.

  Type 2: Only the bachelor's or first professional degree--includes those institutions offering courses of study leading to the customary bachelor of Arts or Bachelor of Science degree and all those degrees which entitle the possessor to enter the profession indicated.

  Type 3: Master's and/or second professional degree--includes those institutions offering the custumary first graduate degree and any degree earned in the same field after the bachelor's or first professional degree. This type of institution does not offer the Doctor of Philosophy or equivalent degrees.

  Type 4: Doctor of Philosophy and equivalent degrees--includes those institutions which are considered to be universities.

- The size stratum for a college was determined by the number of students for which the college reported freshman grades:

  Category 1: 100 students or fewer

Category 2: 101-200 students

Category 3: 201-500 students

Category 4: 501-1000 students

Category 5: 1001 + students

The stratification was accomplished first by sorting the file of colleges on the stratification variables. Then, a random sample of colleges was selected from each size stratum. The sample sizes were chosen to yield a 95% chance that the estimated mean absolute error (defined below) would be within approximately ± .02 of the mean absolute error that would have resulted if all student records had been used. In all, 260 colleges were selected.

The data base of 605 colleges from which the sample was selected consisted of colleges that reported data for 1976 and for at least one of the years 1972, 1973, 1974, and 1975. Therefore not all of the 260 colleges sampled from this data base reported data every year. The number of colleges in the sample, by year, are displayed in Table 1.

Table 1

Summary of Data Base and Sample for Cross-Validation Study

Year	Number of colleges in data base[a]	Number of colleges in sample	Number of student records for Computing Weights	Number of 1976 Student records for Cross-Validation[b]
1972	451	187	97,985	13,732
1973	484	203	114,331	15,589
1974	494	207	108,118	15,412
1975	520	211	109,207	15,384
1976	605	260	----	----

[a]The data base for this study consists of all student records supplied by colleges that participated in ACT's Basic or Standard Research Services in 1976 and in one or more of the years 1972-1975. There are 605 such colleges.

[b]The prediction equations from each year were cross-validated on a random subsample of the 1976 student records.

For each year, about 41% of all colleges which reported data through Basic or Standard were included in the sample.

From each of the years 1972-75, every student record for the sampled colleges was used in computing weights. For 1976, a subsample of the student records was used to cross-validate the prediction equations. The subsampling rate varied according to college size, and was chosen to yield the precision discussed above in estimating mean absolute error. The subsampling rate for 1976 records, by size stratum, are displayed in Table 2.

Table 2

Stratification and Subsampling for the Cross-Validation Study Sample, by Size Stratum

Size Stratum	Number of Colleges in data base	Number of Colleges in sample	Subsampling rate for 1976 records
100 or fewer	47	38	1/2
101-200	221	75	1/4
201-500	202	63	1/4
501-1000	70	44	1/6
Over 1000	65	40	1/1[illegible]
TOTAL	605	260	---

Prediction Equations

Prediction equations were calculated from the 1972, 1973, 1974, and 1975 data using a standard eight variable multiple linear regression:

$$\hat{y} = a_0 + a_1 * \text{ACT English score} + a_2 * \text{ACT Mathematics score} + a_3 * \text{ACT Social Studies score} + a_4 * \text{ACT Natural Science score}$$

$+ a_5$ * high school English grade

$+ a_6$ * high school mathematics grade

$+ a_7$ * high school social studies grade

$+ a_8$ * high school natural science grade

where a_0, a_1, ---, a_8 are regression weights calculated from the base year data.

Prediction equations were also calculated using other techniques, such as high school grades alone, test scores alone, and various ways of combining these two kinds of predictors. In particular, prediction equations were also developed using a technique of the Standard Research Service. In this technique, a prediction based on ACT scores only was averaged with a prediction based on self-reported high school grades. The average is called the TH-index. The decision was made to report results for a standard eight variable multiple linear regression equation because most readers are familiar with this technique. The results with respect to validity over time were virtually identical for either technique.

It is well-known from several studies that males and females differ in the relationship between high school grades, standardized test scores, and freshman grade point average. (See, for example, "Bias in the Uses of Educational Assessment," ACT Technical Report (1973), pp. 229-320, and "Statistical Comparison of Entrance Prediction Equations Using ACT or SAT Scores, or Both," College and University, 1976, Vol. 51, pp. 174-182.) Moreover, colleges often elect to receive separate predictions by sex. Therefore, prediction equations were calculated for males and famales in each college separately, as well as for all students in a college.

Some colleges in the sample submitted student grade data without identifying the sex of their students. Therefore, the sum of the number of males and females from which the separate equations were computed is less than the total number of records from which the overall total group equations were computed.

Cross Validation Statistics

The difference between the predicted and actual grade for a student is called the error of prediction. The magnitude of this error, ignoring the direction, is called the absolute error of prediction. If the absolute error of prediction is averaged over a group of students (for example, all the students in a college) the resulting number is called the mean absolute error (MAE) of prediction in that group of students. This statistic has an immediate relevance to the quality of grade predictions; for example, if the MAE in predicting grade point average in a college is .45, then on the average, there is a discrepancy of .45 (either higher or lower) in predicting grade point average in the college.

A related measure of prediction quality that is sometimes used is the mean squared error, which is the average over a group of students of the squared error of prediction. It is actually this quantity that the usual least squares regression methods seek to minimize. Because the mean squared error is less directly interpretable in practical terms than the MAE, it will not be used here.

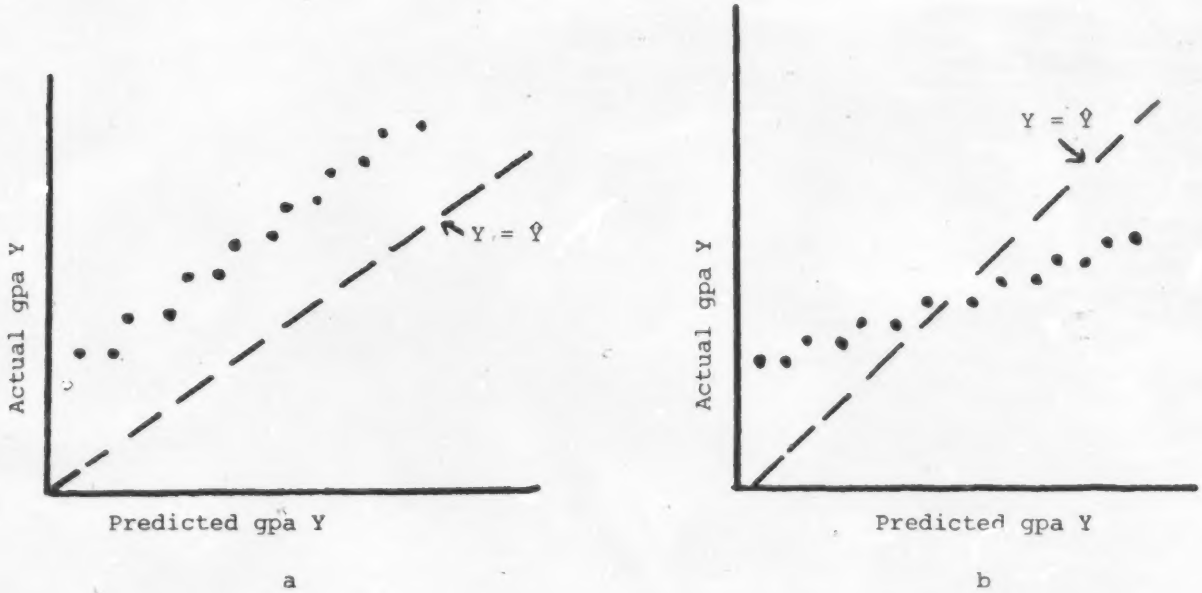
Another useful measure of the quality of prediction for a group of students is the proportion whose predicted grade is within a certain range of the actual grade. In this study we report the proportion of students whose predicted gpa is within 0.20 gpa of the actual gpa. It will be denoted by the abbreviation P20.

Probably the most commonly used measure of predictive efficiency is "cross-validated r," that is, the Pearson correlation between the predicted grades ($\hat{Y}$) and actual grades (Y). Generally, this coefficient is compared with the correlation coefficient calculated from the base year data to give an indication of the stability of the predictions. Because it is so widely used, this statistic is also reported in addition to MAE and P20.

Cross-validated r, being the correlation coefficient between Y and $\hat{Y}$ measures the strength of the linear relationship between them. In practical use, however, this may not be of much interest since the only linear relationship the user cares about is the trivial one $Y = \hat{Y}$. The hypothetical data represented by dots in Figures 1a and 1b illustrate the false conclusions that could result from using cross-validated r as the sole measure of prediction quality. In both cases, r is near 1, yet in both cases the prediction is not very good. In the first instance, the predicted grades are consistently too low; in the second, they are too low some of the time and too high the rest of the time. A large cross-validated r does not guarantee success in prediction. On the other hand, a low value of r will indicate poor prediction, since if Y and $\hat{Y}$ have little linear relationship, they cannot be very close to each other.

Figure 1

Two situations in which cross-validated r
gives a false indication of prediction quality



A justification sometimes advanced for the correlation coefficient is that r^2 measures the proportional decrease in squared error from that which would result if the mean grade were predicted for each student. Similarly, r^2 takes into account the effect on prediction due to the inherent variability of students' grades.

Computations

The prediction equations and cross-validation statistics for each college were computed by a custom program using the IMSL subroutine package.[3] The program was tested on data from three colleges over a three year period. Every computation on these data was compared against one obtained from the SPSS statistical package[4] and where feasible, against one computed by hand. In all cases, no matter what the magnitude of the variable, the results for each college agreed to within ± .001; in a majority of cases the results agreed to within ±.0001. For regression weights and cross-validation statistics, this would indicate close agreement.

The cross-validation statistics from each college were summarized through SPSS routines. The results were weighted to take into account the differential sampling rates in the various strata.

Results

The results of the analyses are displayed in Tables 3, 4, 5, and 6. They indicate that as computed for students as a whole (irrespective of their college), the quality of the prediction equations was quite stable over the four-year period studied. The same is generally true of the quality of the prediction equations for colleges, although a few individual colleges do show some instability.

---

[3] IMSL Library 1 Reference Manual, Edition 6. IMSL, 7500 Bellaire Blvd., Houston, Texas 77036 (1977).

[4] Nie, N.H., et al., SPSS, Statistical Package for the Social Sciences, Second Edition, McGraw Hill, New York (1975).

Table 3 displays the cross-validation statistics MAE, P20, and r as computed for all students in all colleges in the sample. To within sampling error, there is no difference in any of the three statistics over the four year period. MAE for the students is about .53-.55, P20 is about .24-.25, and r, about .55-.56.

Table 3

Results of Cross-Validating Prediction Equations Developed in Years 1972-1975 Against 1976 Freshman Grade Point Average (Results for Students)

Group	Cross-validation statistic*	Year in which prediction equation was developed 72	73	74	75
Total group	MAE	.55	.54	.53	.54
	P20	.24	.24	.25	.25
	r	.55	.55	.56	.56
Males	MAE	.59	.59	.58	.58
	P20	.22	.23	.23	.23
	r	.50	.50	.50	.51
Females	MAE	.53	.52	.51	.52
	P20	.24	.25	.26	.25
	r	.57	.57	.59	n/a

*Note: MAE is the Mean Absolute Error, the average of the unsigned deviations of predicted grades from earned grades.
P20 is the proportion of students whose predicted grade point average was within 0.20 of their earned grade point average.
r is the correlation between predicted and earned grade point average.

Prediction equations developed for males and females separately show the same stability in quality. As expected, the predictions for females were somewhat more accurate (MAE = .51-.53, P20 = .24-.26, r = .57-.59) than the predictions for males (MAE = .58-.59, P20 = .22-.23, r = .50-.51).

Table 4 displays the medians of these cross-validation statistics for colleges. The figures in this table show that measures of prediction quality for students across all colleges are also typical of the measure of prediction quality for individual colleges. Moreover, the same stability in prediction quality over the four years is evident.

Table 4

Results of Cross-Validating Prediction Equations
Developed in Years 1972-75 Against 1976 Freshman Grade Point Average
(Median Cross-Validation Statistics for Colleges)

Group	Cross-validation statistic*	Year in which prediction equation was developed			
		72	73	74	75
Total group	MAE	.53	.53	.52	.51
	P20	.24	.24	.25	.25
	r	.54	.55	.56	.56
Males	MAE	.57	.60	.57	.56
	P20	.21	.21	.21	.23
	r	.52	.51	.51	.53
Females	MAE	.52	.51	.50	.52
	P20	.25	.25	.26	.25
	r	.57	.57	.59	.58

*Note: MAE is the Mean Absolute Error, the average of the unsigned deviations of predicted grades from earned grades.
P20 is the proportion of students whose predicted grade point average was within 0.20 of their earned grade point average.
r is the correlation between predicted and earned grade point average.

Table 5 is a frequency distribution of MAE among colleges for the four years studied. For about 72%-78% of the colleges, the MAE is 0.60 or less. Thus, the MAE reported for students over all colleges is fairly typical of the MAE's in most colleges individually. It should be noted, though, that for a small proportion (2%-4%) of colleges, MAE's in the range 0.80-1.10 occurred. No college in the sample had a MAE higher than 1.10.

Table 5

Proportions of Colleges with Various Ranges of MAE

Range in MAE*	Year			
	72	73	74	75
0.0 - 0.1	.00	.00	.00	.00
0.1 - 0.2	.00	.00	.00	.00
0.2 - 0.3	.00	.00	.00	.00
0.3 - 0.4	.07	.08	.12	.09
0.4 - 0.5	.31	.31	.29	.36
0.5 - 0.6	.34	.34	.37	.33
0.6 - 0.7	.18	.21	.15	.13
0.7 - 0.8	.09	.05	.06	.05
0.8 - 0.9	.02	.02	.01	.02
0.9 - 1.0	.00	.00	.00	.01
1.0 - 1.1	.00	.00	.00	.01

*Note: The ranges in MAE are defined so that, for example, the fourth range includes colleges with MAE greater than or equal to 0.3, but less than 0.4.

The relative stability of the quality of predictions over all students and colleges could mask instability in particular colleges. Table 6, which addresses this issue, is a frequency distribution of the differences between MAE for equations developed in 1975 and MAE for equations developed in 1974, 1973, and 1972. The results indicate that for most colleges, MAE is quite stable from year to year. About 56% of the colleges have a greater MAE from predictions developed in 1974 then in 1975; this is nearly balanced by the proportion of colleges with a smaller MAE from predictions developed in 1974 than in 1975 (44%). Moreover, about 99% of the differences in MAE are in the interval (-.10,.10). Slightly more colleges have a larger MAE from

equations developed in 1974, 1973, or 1972 than 1975, but the differences in MAE are concentrated in the range of -.10 to .00. In summary, the difference in MAE for predictions developed in 1975 and predictions developed in 1972, 1973, or 1974 is between -.10 and +.10 for over 90% of the colleges

Table 6

Proportions of Colleges with Various Differences between MAE for 1975 and MAE for 1972-1974

Range of Difference	MAE(1975)-MAE(1972)	MAE(1975)-MAE(1973)	MAE(1975)-MAE(1974)
-0.50 to -0.40	.00	.00	.00
-0.40 to -0.30	.00	.00	.00
-0.30 to -0.20	.00	.00	.00
-0.20 to -0.10	.06	.03	.01
-0.10 to 0.00	.63	.69	.55
0.00 to 0.10	.30	.28	.44
0.10 to 0.20	.00	.00	.00
0.20 to 0.30	.01	.00	.01
0.30 to 0.40	.00	.01	.00
0.40 to 0.50	.00	.00	.00
0.50 to 0.60	.00	.00	.00

Summary and Conclusions

Prediction equations based on ACT scores and high school grades were calculated for 260 randomly sampled colleges that participated in the ACT Research Services during the period 1972-1976. Separate prediction equations were calculated from data for the years 1972, 1973, 1974, and 1975. The prediction equations for each year were then cross-validated against freshman grade point averages earned in 1976.

The results indicate that for most students and colleges, the quality of the prediction equations is quite stable over the four years. The mean absolute error of prediction ranged from .53 to .55, the proportion of students whose predicted grade was within .20 of their earned grade ranged from .24 to .25, and the cross-validated correlation ranged from .55 to .56. A similar stability was noted in the accuracy of grade predictions based on males and females separately.

One can conclude, therefore, that although most colleges experience some change in their students' academic abilities, their curricula, or grading practices over a period of time, freshman gpa can in most cases be reliably predicted using equations as old as four years.

The results of the study also suggest (but do not confirm) that colleges may pool data from two or more years in developing prediction equations. Pooling is done by colleges with small student populations to increase the amount of data (and therefore, the accuracy) in developing prediction equations.